理科教科内容指導論Ｉ：物理分野

物理現象の定量的把握
第4回目

(実験)データの眺め方 ~ 統計学の基礎 続き

統計のはなし―基礎・応用・娯楽 (Best selected business books)

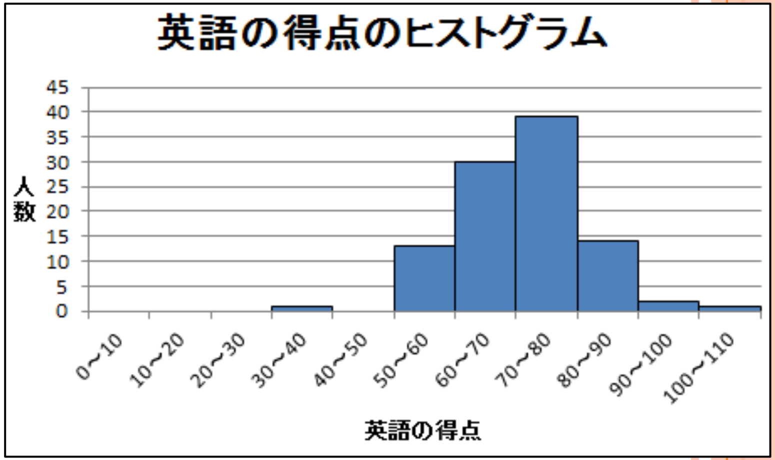
大村 平

日科技連出版社

1836円

前回の復習と今回以降の目標

東京女子大学 小西善二郎氏 Webサイトより



- データ
- ヒストグラム
- 代表値 (平均値、最頻値、中間値)
- 分布の散らばり

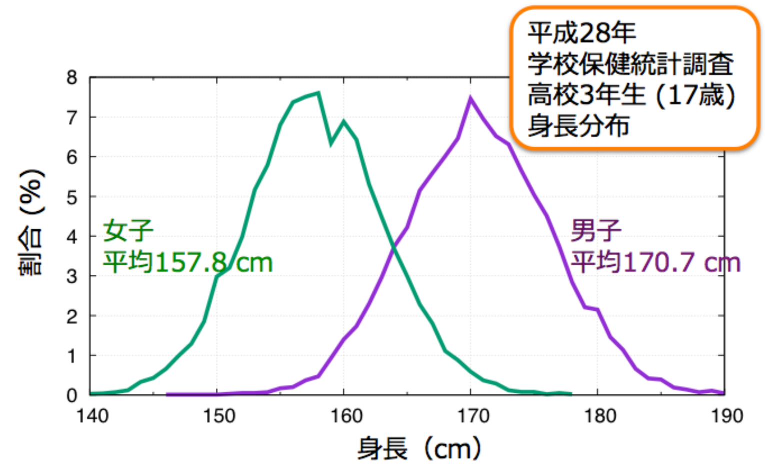
- 「集団の分布」(ヒストグラム) が分かれば代表値や散らばりは計算できる。

→「集団の分布」のパターンを理解することが重要。

- 「集団の分布」のパターンの重要な一つである"正規分布"を理解。

前回の復習と今回以降の目標

- データ
- ヒストグラム
- 代表値 (平均値、最頻値、中間値)
- 分布の散らばり



- 「集団の分布」(ヒストグラム) が分かれば代表値や散らばりは計算できる。

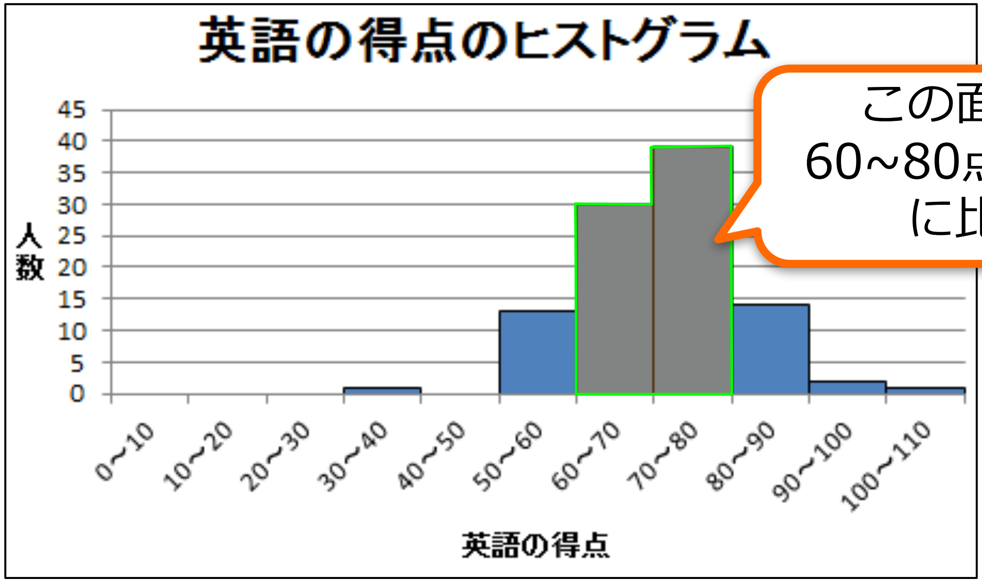
→「集団の分布」のパターンを理解することが重要。

- 「集団の分布」のパターンの重要な一つである"正規分布"を理解する。

集団の分布のパターンの表し方

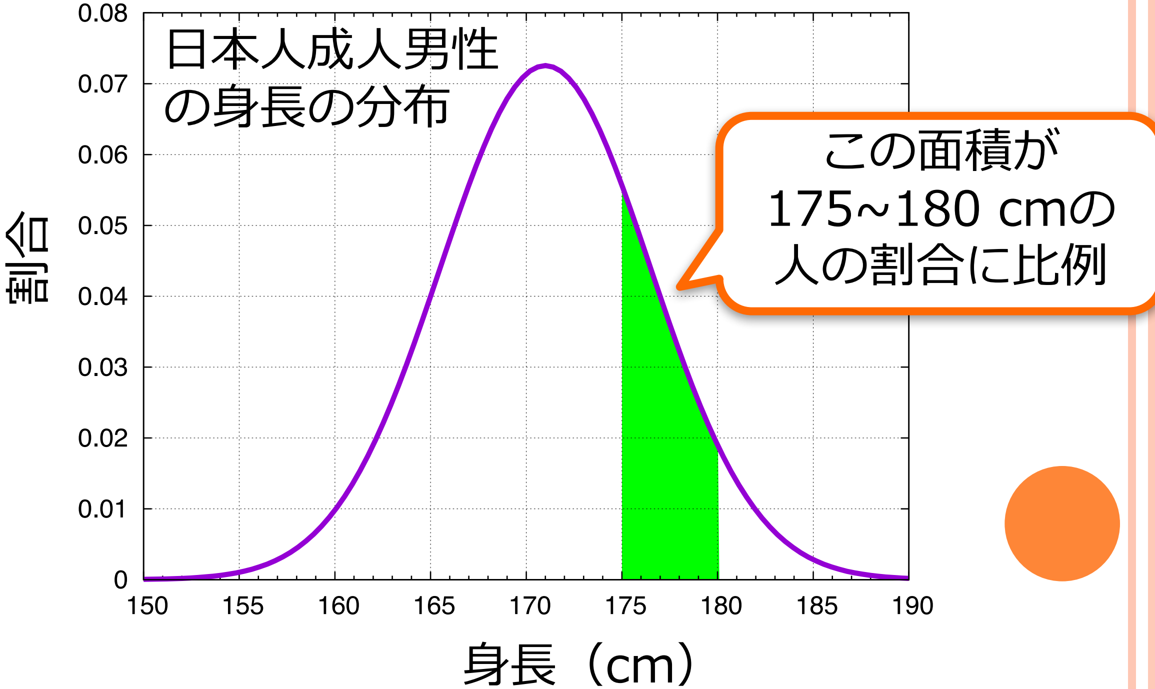
- ヒストグラムを使えば良い。

→ヒストグラムの**面積**は、その値の範囲に属するデータの数と比例する。



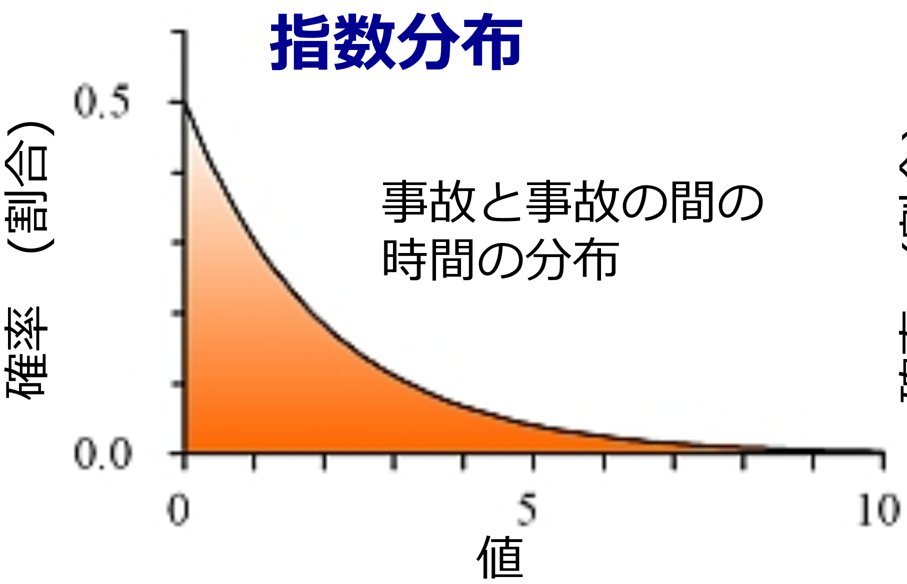
- この考え方を連続的な数にも拡張したものを確率密度分布と呼ぶ。

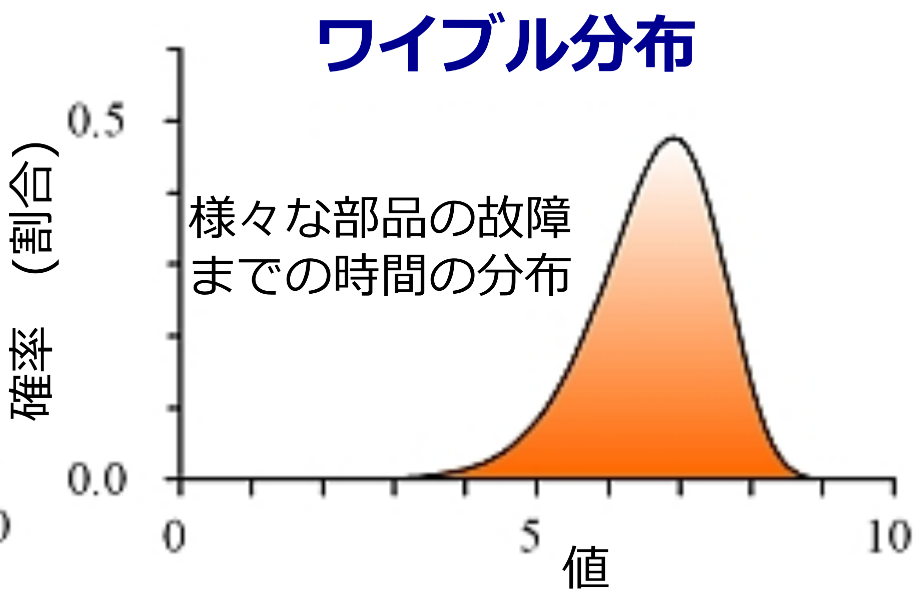
→面積がその値の範囲に属するデータの**割合**と比例する。



様々な分布

- 集団の分布 → **割合**の分布 → **確率**の分布 (確率分布)
- 分布のパターンは様々。



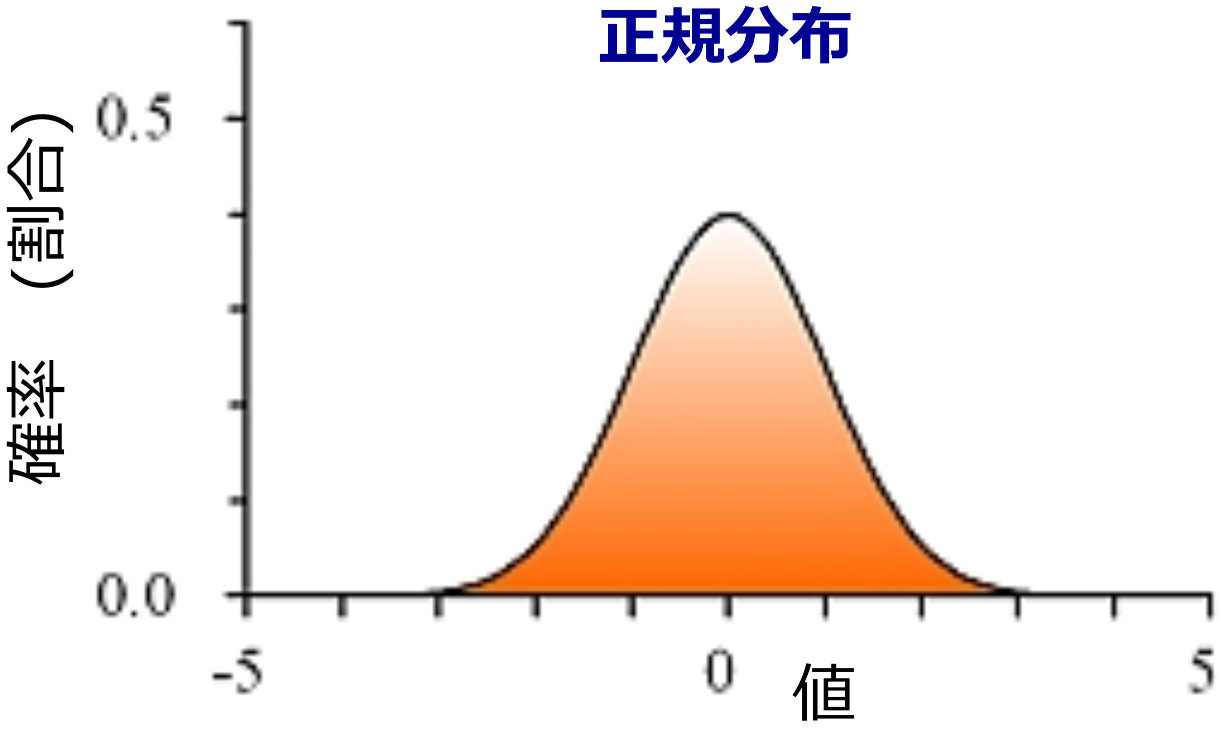


http://www.biwako.shiga-u.ac.jp/sensei/mnaka/ut/statdist.html

- 一様分布、パレート分布、コーシー分布、超幾何分布 etc…
  - 興味があれば統計学の教科書 (「統計学入門」など)を読んでみいて下さい。案外面白いです。

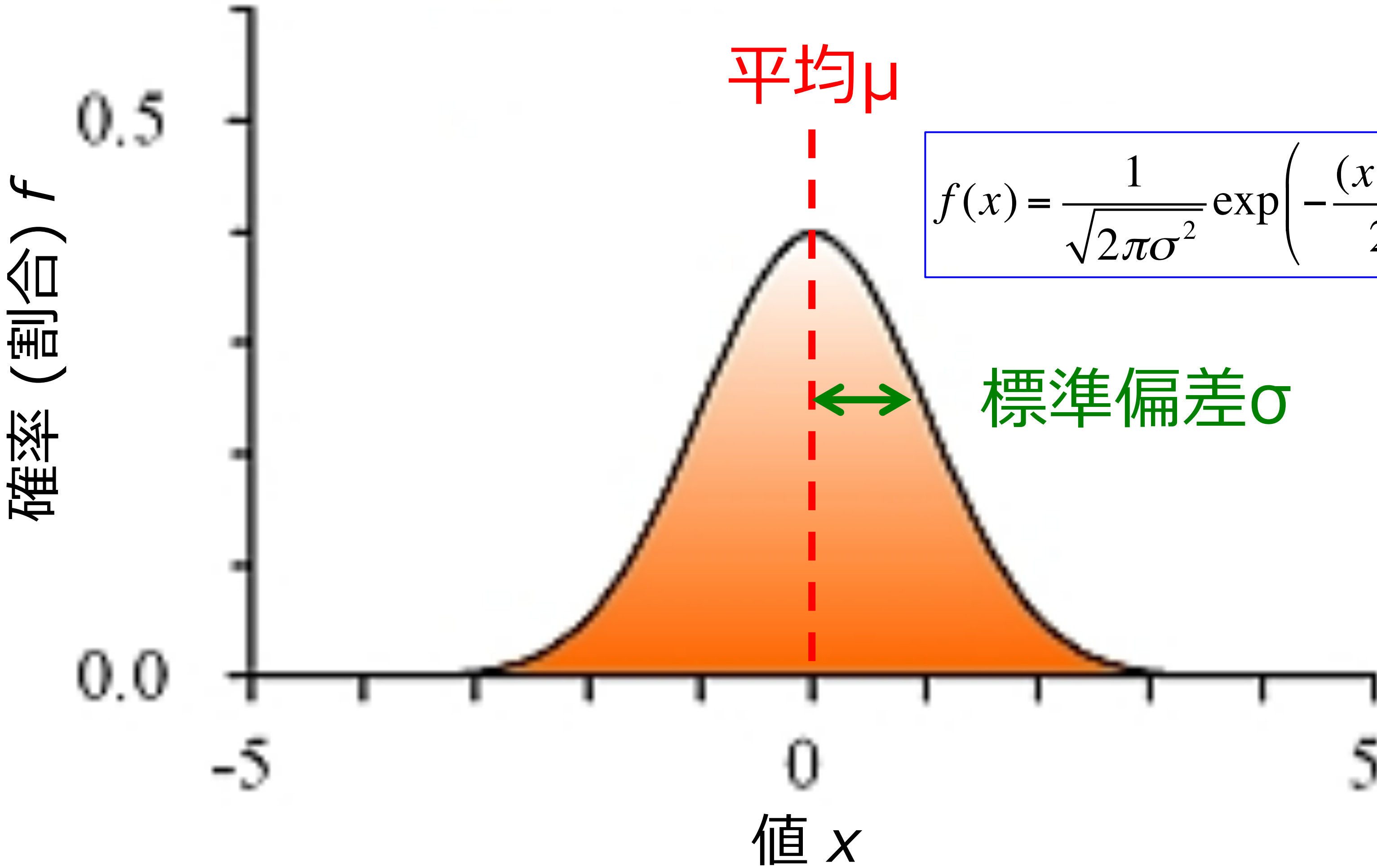
正規分布

- 無限に広がる、釣鐘状の左右対称な分布。
- **<u>最頻値</u>**と**<u>中間値</u>**と**<u>平均値</u>**が一致した素直な分布。
- 色々な物がこの分布に従うことが知られている。
  - 生物の身長・体重、試験の点数 (厳密には違う)etc.
  - 測定の誤差 (誤差分布)
  - 平均値の分布 ... 中心極限定理、大数の法則



正規分布

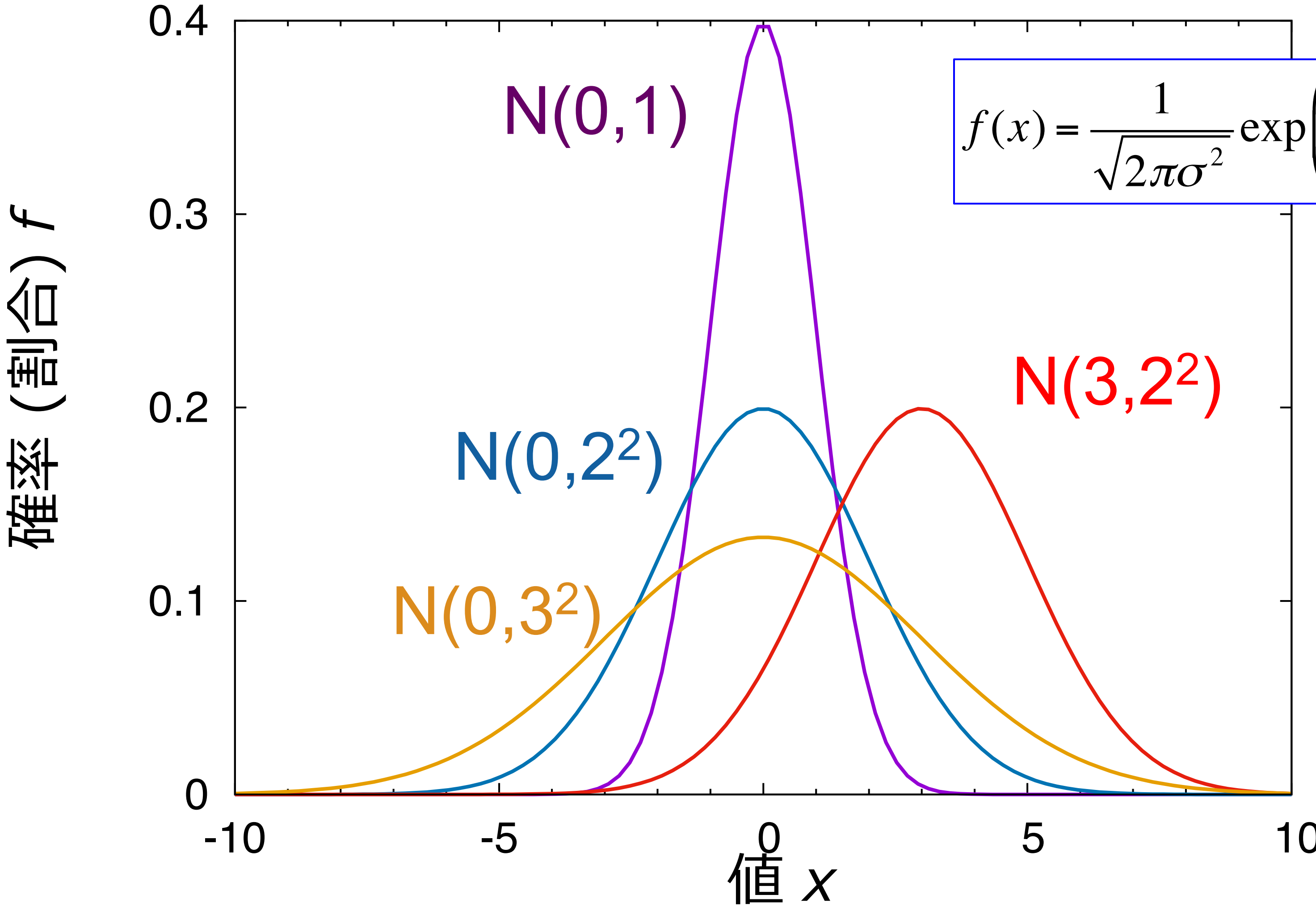
- 平均μ (ミュー) と標準偏差σ (シグマ)の値を決めると分布の形が決まる。
- 平均μ、標準偏差σの 正規分布をN(μ, σ²)と表記する。



$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

正規分布

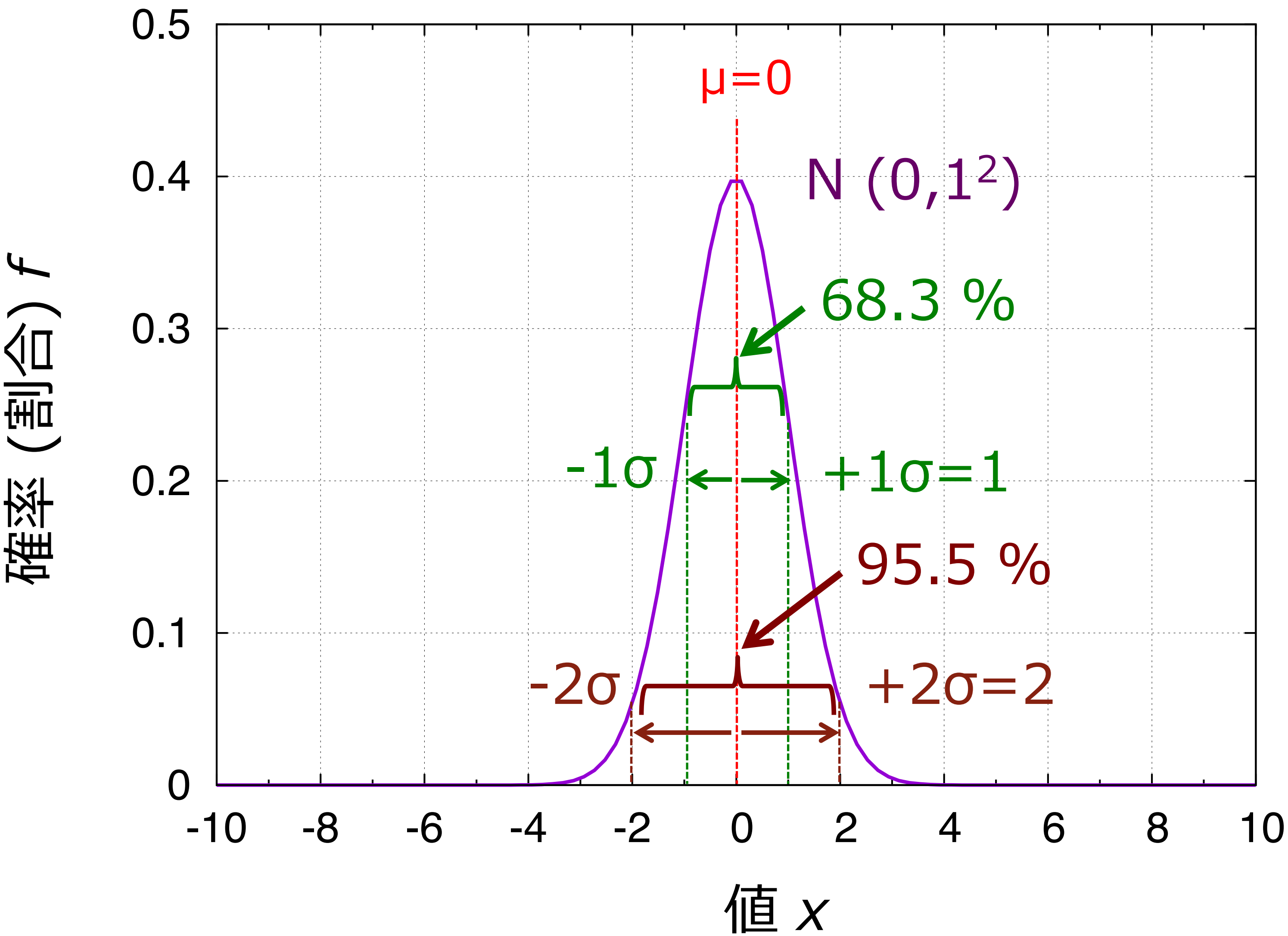
- 平均μ (ミュー) と標準偏差σ (シグマ)の値を決めると分布の形が決まる。
- 平均μ、標準偏差σの 正規分布を**N(μ, σ²)**と表記する。



$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right)$$

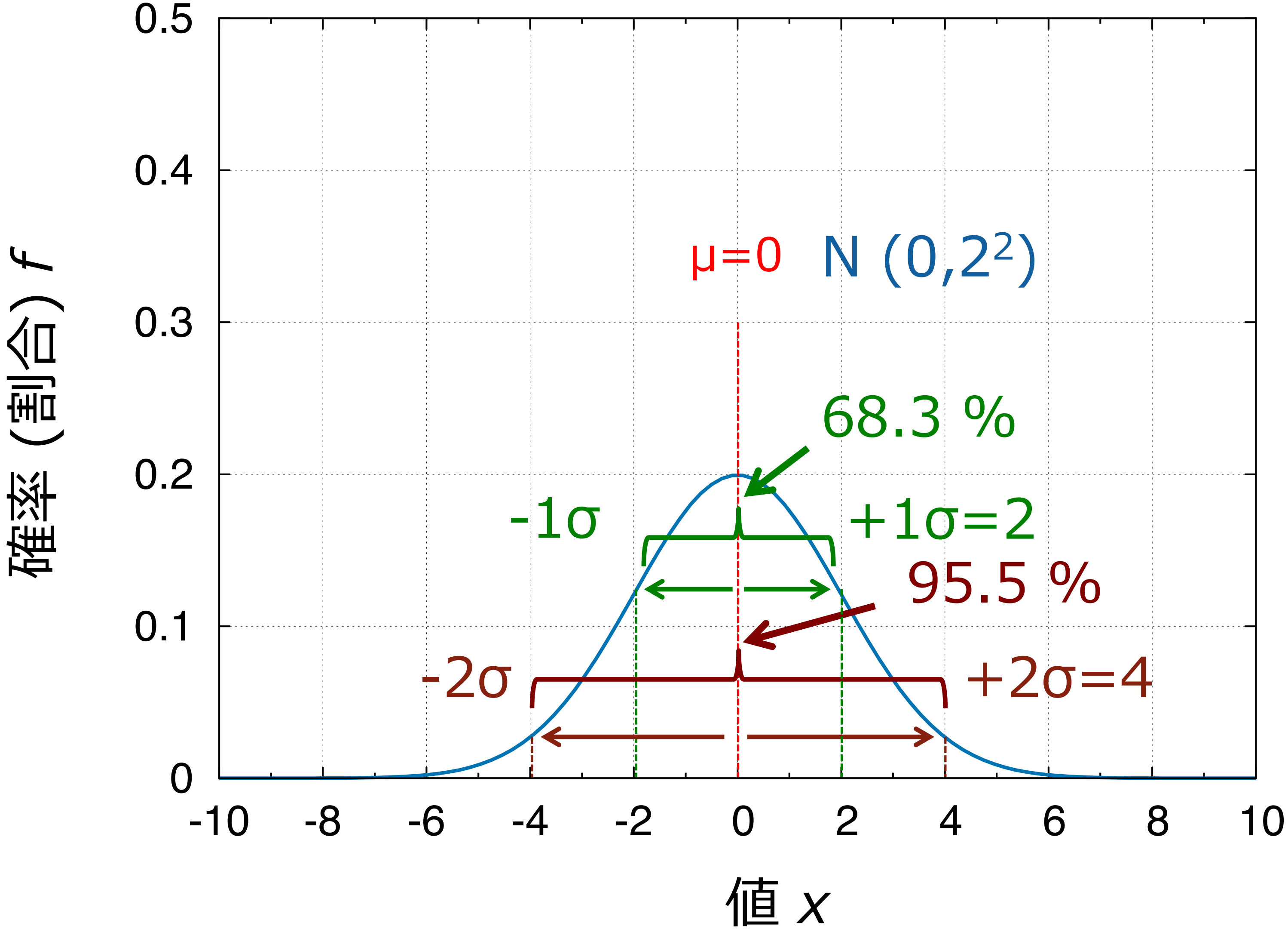
正規分布の特徴

- 平均μから標準偏差σを単位としてある幅をとると、その範囲の<u>**面積**</u>がどんな正規分布N(μ,σ²)の場合にも等しい値になる。



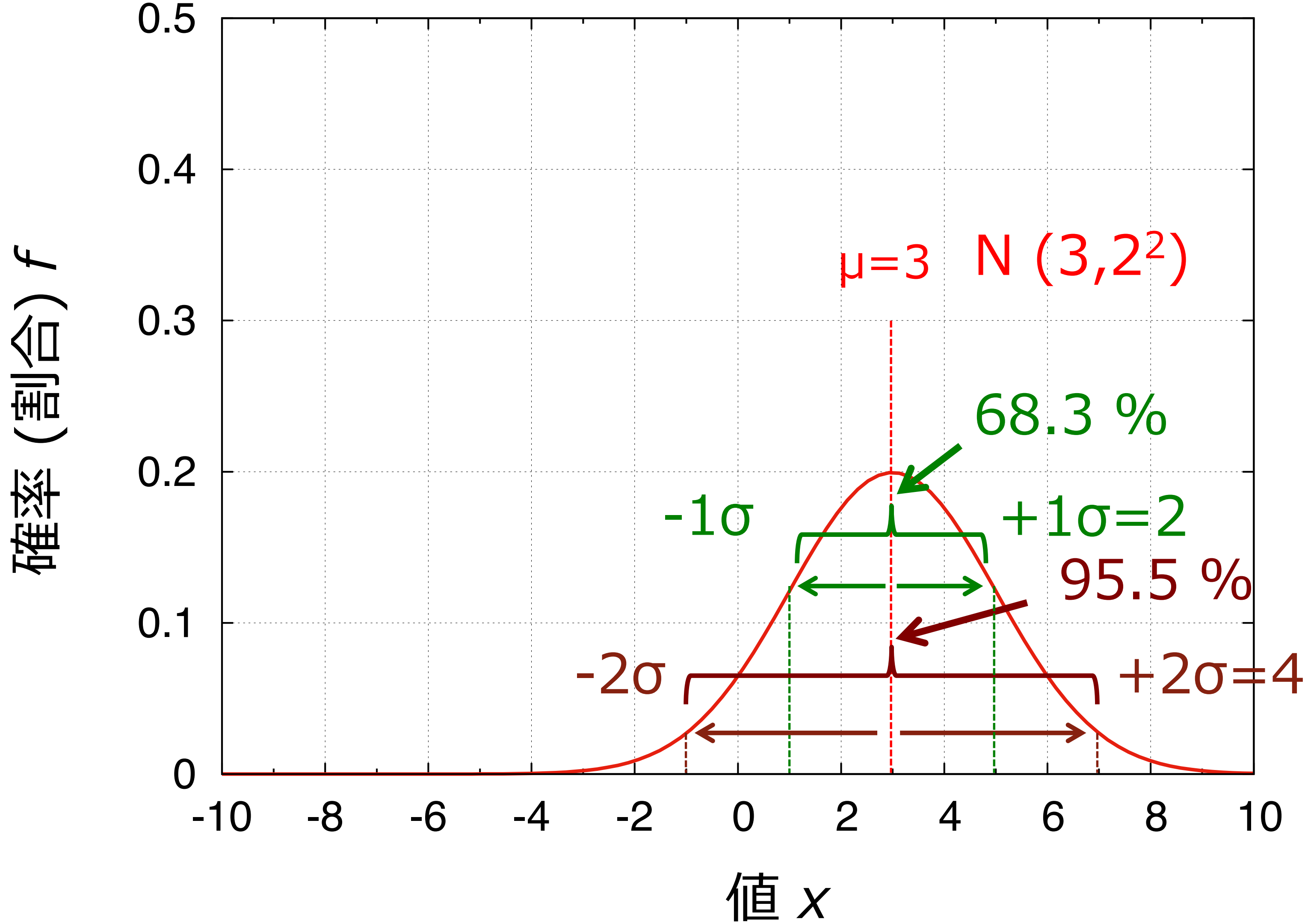
正規分布の特徴

- 平均μから標準偏差σを単位としてある幅をとると、その範囲の面積がどんな正規分布N(μ,σ²)の場合にも等しい値になる。



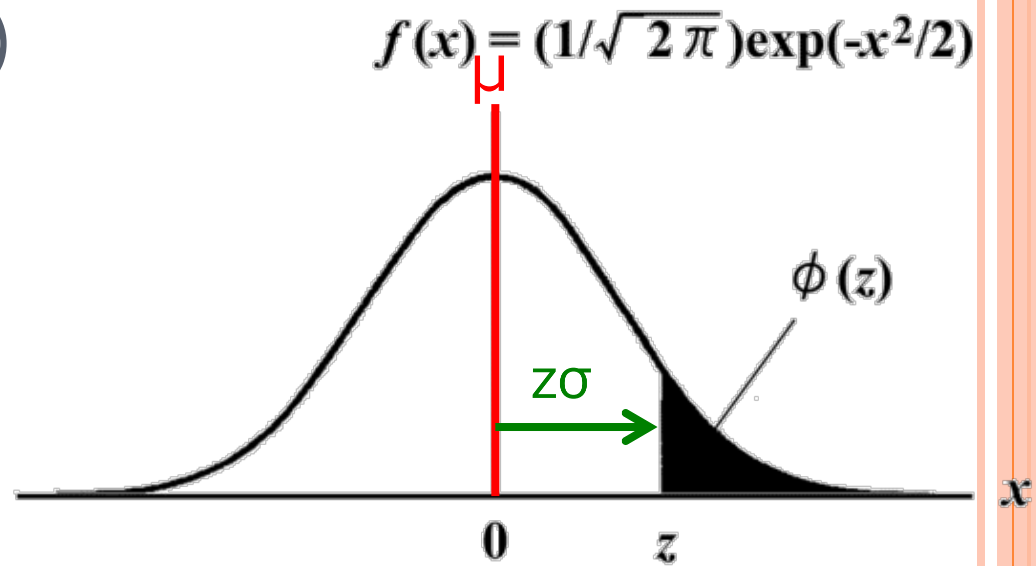
正規分布の特徴

- 平均μから標準偏差σを単位としてある幅をとると、その範囲の面積がどんな正規分布$N(\mu,\sigma^2)$の場合にも等しい値になる。



https://staff.aist.go.jp/t.ihara/normsdist.html

正規分布表 (上側確率)



- 他に表の流儀は幾つかあります。
- EXCELでも

”1－NORMSDIST(z)”

で計算可能。

- 例えばμ+1.24σ 以上が全体に占める割合は10.7%。

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	5.00E-01	4.96E-01	4.92E-01	4.88E-01	4.84E-01	4.80E-01	4.76E-01	4.72E-01	4.68E-01	4.64E-01
0.1	4.60E-01	4.56E-01	4.52E-01	4.48E-01	4.44E-01	4.40E-01	4.36E-01	4.33E-01	4.29E-01	4.25E-01
0.2	4.21E-01	4.17E-01	4.13E-01	4.09E-01	4.05E-01	4.01E-01	3.97E-01	3.94E-01	3.90E-01	3.86E-01
0.3	3.82E-01	3.78E-01	3.74E-01	3.71E-01	3.67E-01	3.63E-01	3.59E-01	3.56E-01	3.52E-01	3.48E-01
0.4	3.45E-01	3.41E-01	3.37E-01	3.34E-01	3.30E-01	3.26E-01	3.23E-01	3.19E-01	3.16E-01	3.12E-01
0.5	3.09E-01	3.05E-01	3.02E-01	2.98E-01	2.95E-01	2.91E-01	2.88E-01	2.84E-01	2.81E-01	2.78E-01
0.6	2.74E-01	2.71E-01	2.68E-01	2.64E-01	2.61E-01	2.58E-01	2.55E-01	2.51E-01	2.48E-01	2.45E-01
0.7	2.42E-01	2.39E-01	2.36E-01	2.33E-01	2.30E-01	2.27E-01	2.24E-01	2.21E-01	2.18E-01	2.15E-01
0.8	2.12E-01	2.09E-01	2.06E-01	2.03E-01	2.00E-01	1.98E-01	1.95E-01	1.92E-01	1.89E-01	1.87E-01
0.9	1.84E-01	1.81E-01	1.79E-01	1.76E-01	1.74E-01	1.71E-01	1.69E-01	1.66E-01	1.64E-01	1.61E-01
1.0	1.59E-01	1.56E-01	1.54E-01	1.52E-01	1.49E-01	1.47E-01	1.45E-01	1.42E-01	1.40E-01	1.38E-01
1.1	1.36E-01	1.33E-01	1.31E-01	1.29E-01	1.27E-01	1.25E-01	1.23E-01	1.21E-01	1.19E-01	1.17E-01
1.2	1.15E-01	1.13E-01	1.11E-01	1.09E-01	1.07E-01	1.06E-01	1.04E-01	1.02E-01	1.00E-01	9.85E-02
1.3	9.68E-02	9.51E-02	9.34E-02	9.18E-02	9.01E-02	8.85E-02	8.69E-02	8.53E-02	8.38E-02	8.23E-02

(例題)松坂大輔と大谷翔平

- 「松坂・大谷は日本人男性の中ではどれくらい背が高い方か?」
- 「彼らより背の高い日本人男性は全体の何%いるか?」
- 日本人男性の身長の分布は正規分布だと仮定して計算してみる。
- H25年度 30~34歳日本人男性
  - 平均身長μ 172.5 cm
  - 標準偏差σ 5.5 cm

(総務省 統計局 https://www.e-stat.go.jp/SG1/estat/GL02020101.domethod=xlsDownload&fileId=000007744784&releaseCount=2)

183.0 cm

193.0 cm

Wikipedia

(例題)松坂大輔と大谷翔平

- H25年度 30~34歳日本人男性
  - 平均身長μ 172.5 cm
  - 標準偏差σ　5.5 cm

(総務省 統計局 http://www.e-stat.go.jp/SG1/estat/Xlsdl.do?sinfid=000027237853)

- 松坂大輔
- 平均からズレ
    - 183.0-172.5=10.5 cm
  - 標準偏差でこのズレを表すと
    - 10.5/5.5 = 1.91 σ
  - 1.91σ以上が占める割合は
    - 2.81 %
    - $1/(2.81\times10^{-2})$=36人
  - 「松坂より背の高い日本人男性は全体の2.81 %で、36人に1人くらい。」

183.0 cm

193.0 cm

Wikipedia

(例題)松坂大輔と大谷翔平

- H25年度 30~34歳日本人男性
  - 平均身長μ 172.5 cm
  - 標準偏差σ 5.5 cm

(総務省 統計局 http://www.e-stat.go.jp/SG1/estat/Xlsdl.do?sinfid=000027237853)

- 大谷翔平

183.0 cm

193.0 cm

Wikipedia

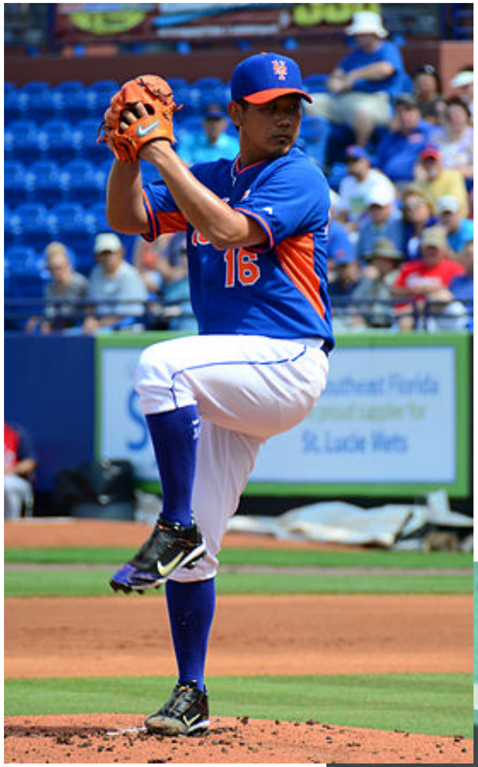
- 「大谷より背の高い日本人男性は全体の________________%で、________________人に1人くらい。」

(例題)松坂大輔と大谷翔平

- H25年度 30~34歳日本人男性
  - 平均身長μ 172.5 cm
  - 標準偏差σ　5.5 cm

(総務省 統計局 http://www.e-stat.go.jp/SG1/estat/Xlsdl.do?sinfid=000027237853)

- 大谷翔平
  - 平均からズレ
    - 193.0-172.5=20.5 cm
  - 標準偏差でこのズレを表すと
    - 20.5/5.5 = 3.73 σ
  - 3.73σ以上が占める割合は
    - 9.57×10^{-3} %
    - $1/(9.57\times10^{-5})=1.04\times10^{4}$人
  - 「大谷より背の高い日本人男性は全体の9.57×10^{-3} %で、1.04×10^{4}人に1人くらい。」



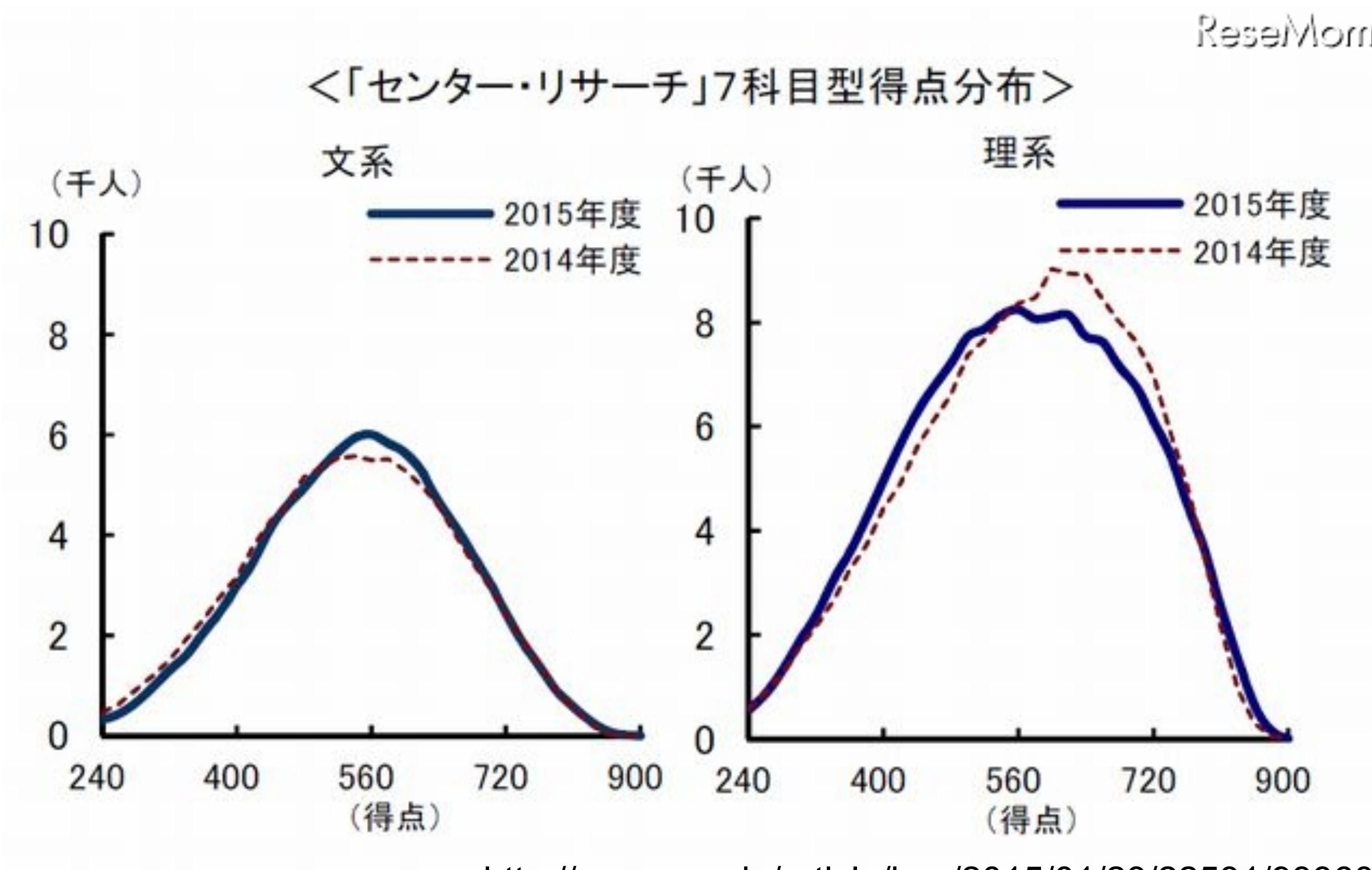
183.0 cm

193.0 cm

Wikipedia

(例題)東大 理IIIはどのくらい凄いのか?

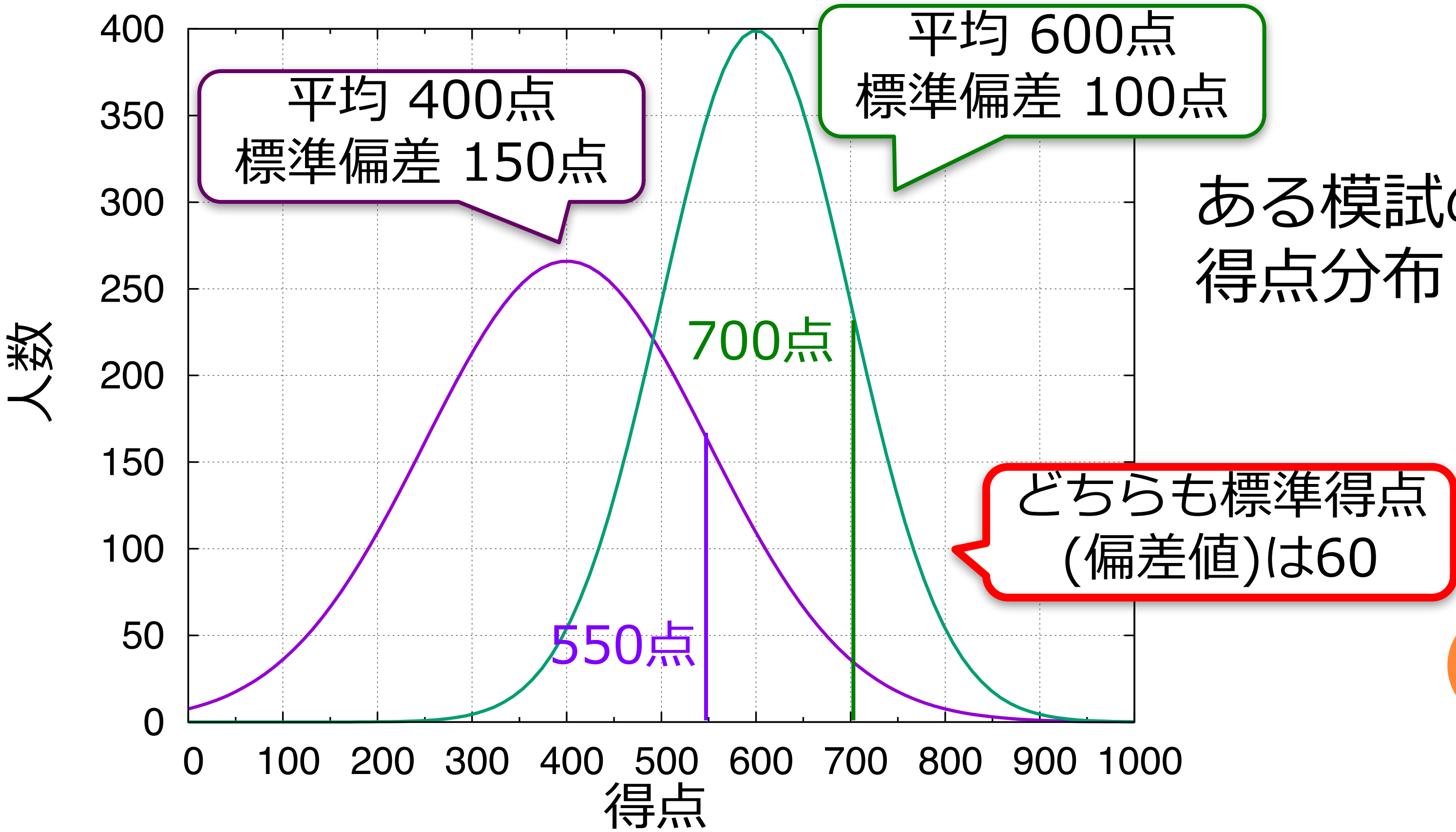
- 受験生の学力 (得点) は正規分布だと仮定する。
  - この仮定はあまり正しくないです。



http://resemom.jp/article/img/2015/01/23/22531/93360.html

標準得点 (偏差値)

- 標準得点 $$Z_i = 50 + 10\left(\frac{x_i - \bar{x}}{\sigma}\right)$$
  - 平均や標準偏差の異なる集団に属するデータを「集団の中でどの辺りに位置するか」で表すことで、互いに比較できるようにする。



ある模試の得点分布

(例題)東大 理IIIはどのくらい凄いのか?

- 受験生の学力 (得点) は正規分布だと仮定する。
  - この仮定はあまり正しくないです。

- “偏差値” (標準得点) は平均50、標準偏差10にした得点。
  - $N(50, 10^2)$の分布。$\mu=50$ $\sigma=10$

- 2016年度 偏差値 74.8
  - http://daigakujuken-plus.com/nyuushi-hensati-ranking/toudai/rika3rui/

- 「東大 理IIIに入るには全受験生の上位________%以上に入る必要がある。」

(例題)東大 理IIIはどのくらい凄いのか?

- 受験生の学力 (得点) は正規分布だと仮定する。
  - この仮定はあまり正しくないです。
- “偏差値” (標準得点) は平均50、標準偏差10にした得点。
  - $N(50, 10^2)$の分布。$\mu=50$ $\sigma=10$
- 2016年度 偏差値 74.8 → 2.48σ
  - http://daigakujuken-plus.com/nyuushi-hensati-ranking/toudai/rika3rui/
- 2.48σ以上の占める割合は0.657%
- 「東大 理IIIに入るには全受験生の上位0.657%以上に入る必要がある。」
  - “希少性”という意味では身長186.1 cm (=172.5+2.48×5.5) 以上の男性と同じ。

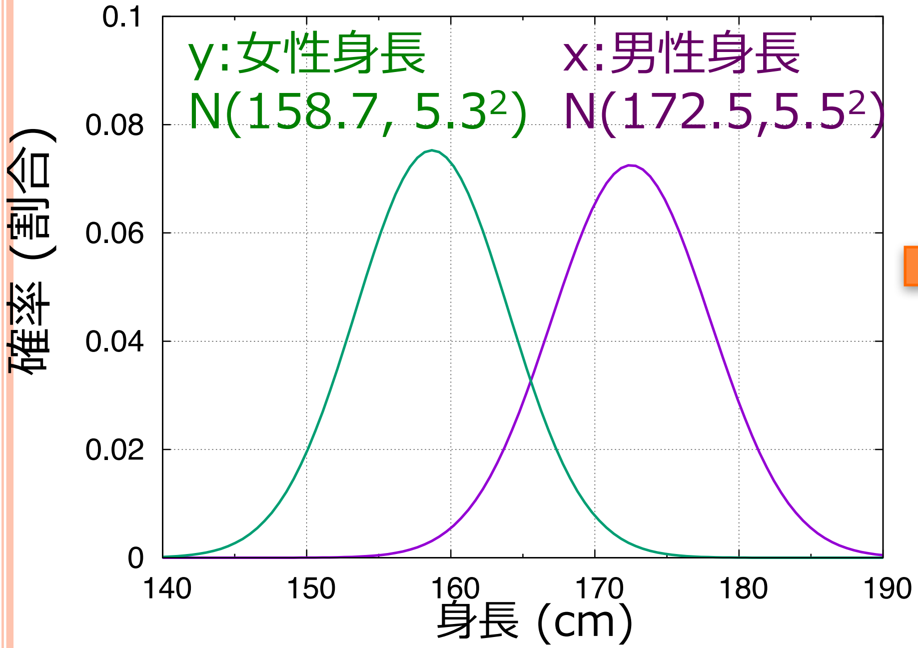
正規分布の加法性

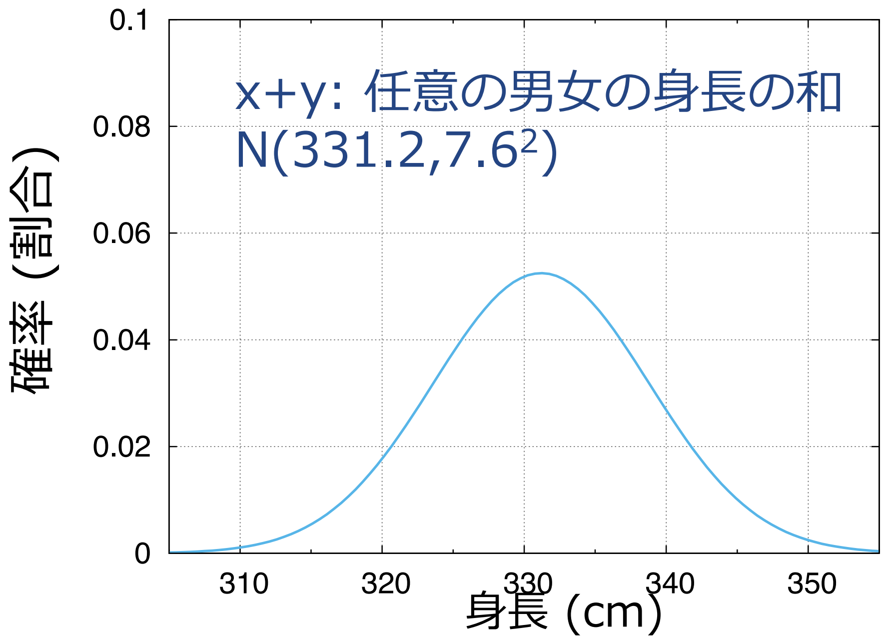
- $N(\mu_1,\sigma_1{}^2)$の分布のxと、$N(\mu_2,\sigma_2{}^2)$の分布のyがある時、ax+byの分布は$N(\underline{\mathbf{a\mu_1+b\mu_2}}, \underline{\mathbf{(a\sigma_1)^2+(b\sigma_2)^2}})$となる。(正規分布の加法性)

- 日本人男性成人身長の分布 N(172.5 cm, $(5.5\ \mathrm{cm})^2$)
- 日本人女性成人身長の分布 N(158.7 cm, $(5.3\ \mathrm{cm})^2$)
  - H25年度 30~34歳 (総務省 統計局 http://www.e-stat.go.jp/SG1/estat/Xlsdl.do?sinfid=000027237853)
- 無作為に日本人成人男女2名を選び出すと、その身長の和の分布は
  $N(172.5+158.7\ \mathrm{cm}, 5.5^2+5.3^2\ \mathrm{cm}^2)$
  $=N(331.2\ \mathrm{cm}, 58.3\ \mathrm{cm}^2)$
  $=N(331.2\ \mathrm{cm}, (7.6\ \mathrm{cm})^2)$
  - 平均 $\mu=331.2$ cm 標準偏差 $\sigma=(58.3)^{0.5}=7.6$ cm

正規分布の加法性

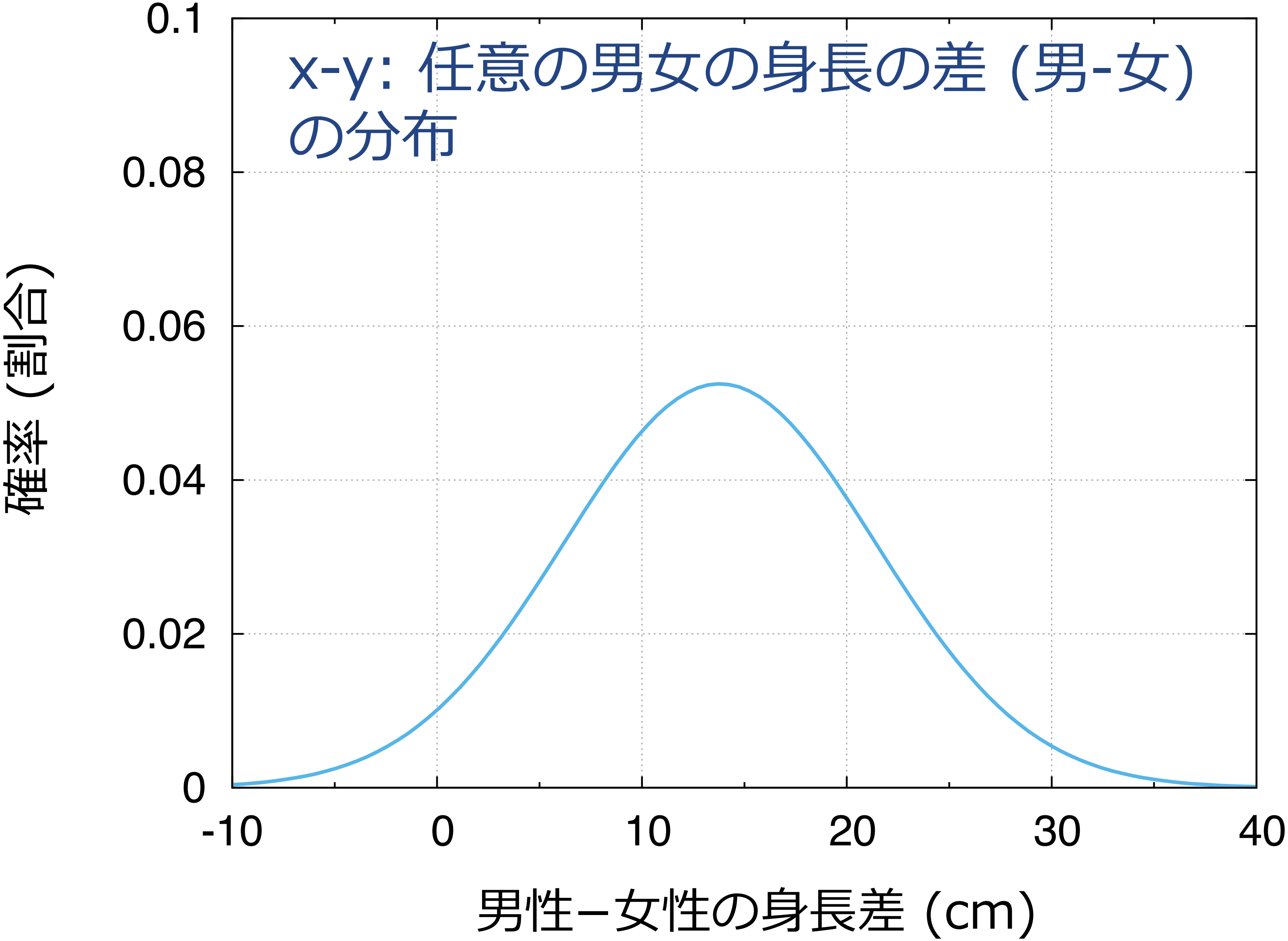
- $N(\mu_1,\sigma_1^2)$の分布のxと、$N(\mu_2,\sigma_2^2)$の分布のyがある時、
  ax+byの分布は$N(a\mu_1+b\mu_2, (a\sigma_1)^2+(b\sigma_2)^2)$となる。
  (正規分布の加法性)

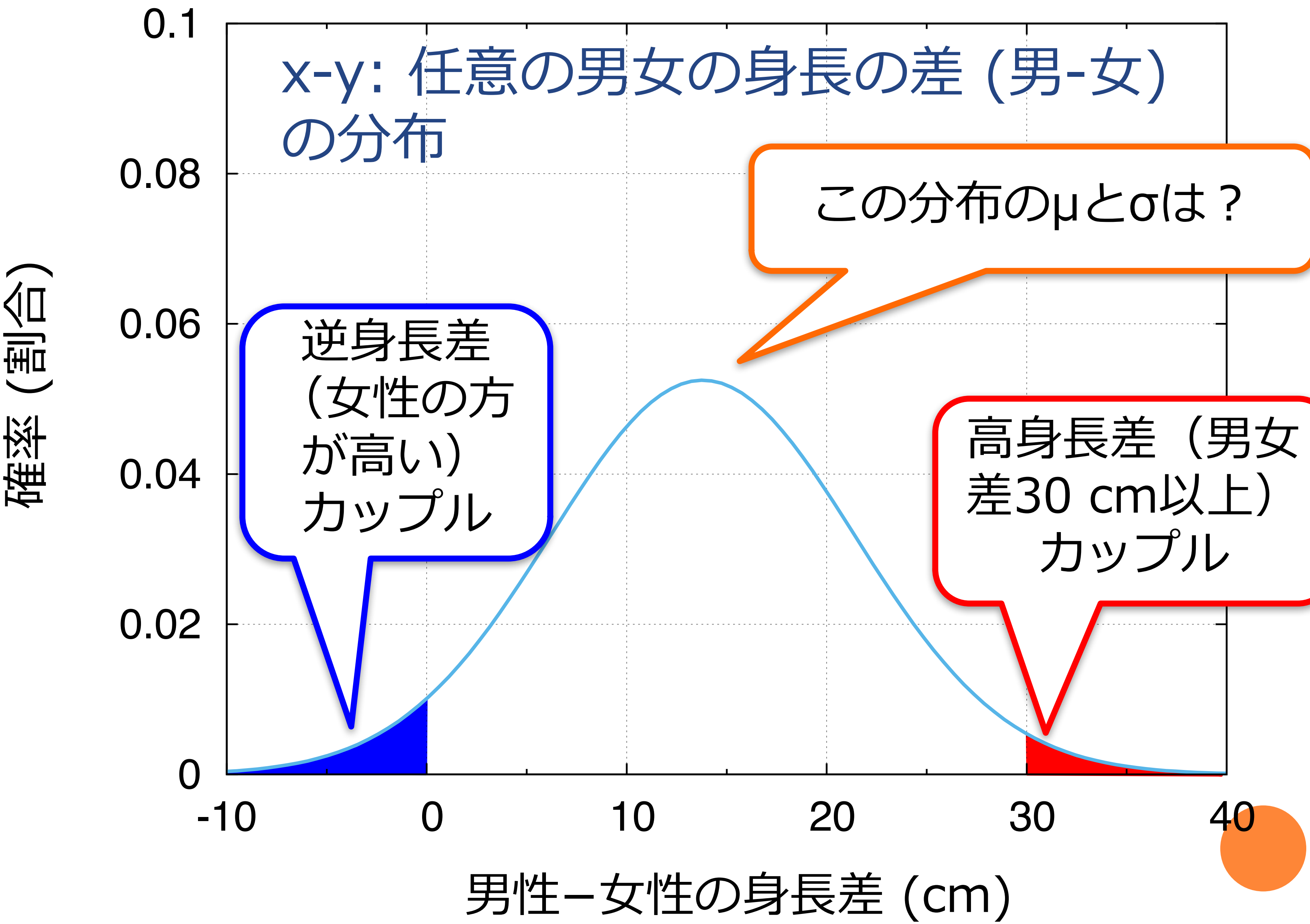




正規分布の加法性

- $N(\mu_1,\sigma_1{}^2)$の分布のxと、$N(\mu_2,\sigma_2{}^2)$の分布のyがある時、ax+byの分布は$N(a\mu_1+b\mu_2, (a\sigma_1)^2+(b\sigma_2)^2)$となる。(正規分布の加法性)
- 日本人男性成人身長の分布 N(172.5 cm, (5.5 cm)^2)
- 日本人女性成人身長の分布 N(158.7 cm, (5.3 cm)^2)
  - H25年度 30~34歳 (総務省 統計局 http://www.e-stat.go.jp/SG1/estat/XlsdI.do?sinfid=000027237853)
- 無作為に日本人成人男女2名を選ぶと、その身長の差の分布はN(____________cm, (____________cm)^2)
- もし世の中の男女カップルが身長差とは無作為に相手を選んでいるとしたら
  - 逆身長差カップル(女性の方が高い) は全体の________%
  - 高身長差カップル(男性の方が30 cm以上高い)は全体の________%


x-y: 任意の男女の身長の差 (男-女)
の分布
確率（割合）
0.1
0.08
0.06
0.04
0.02
0
-10
0
10
20
30
40
男性－女性の身長差 (cm)


x-y: 任意の男女の身長の差 (男-女)の分布
この分布のμとσは？
逆身長差（女性の方が高い）カップル
高身長差（男女差30 cm以上）カップル
0.1
0.08
0.06
0.04
0.02
0
-10
0
10
20
30
40
確率（割合）
男性−女性の身長差 (cm)

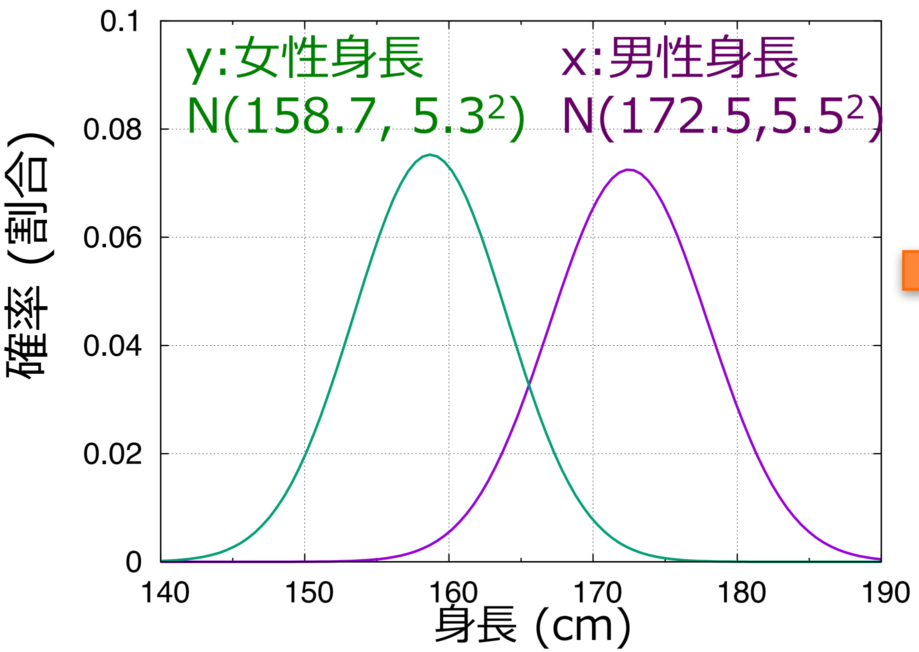
正規分布の加法性

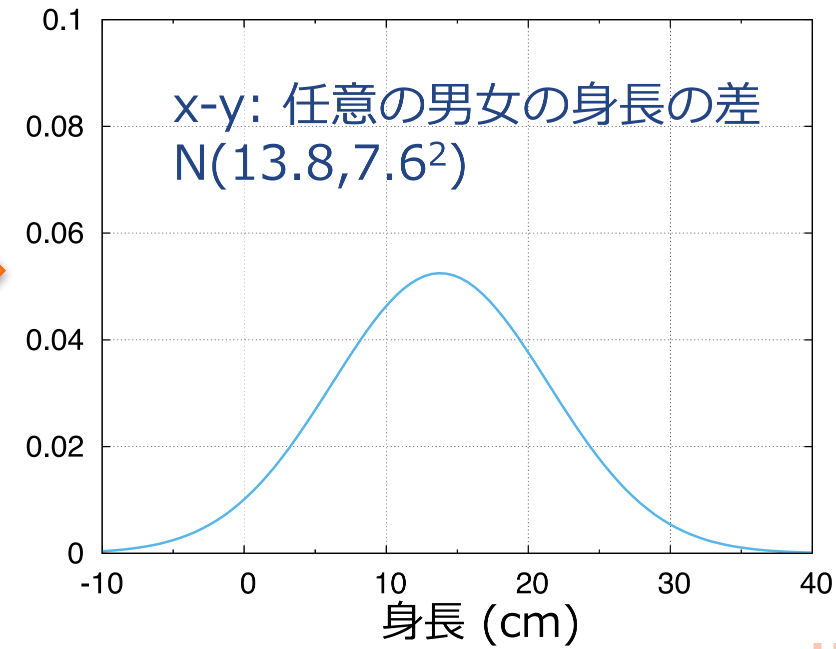
- $N(\mu_1,\sigma_1^2)$の分布のxと、$N(\mu_2,\sigma_2^2)$の分布のyがある時、ax+byの分布は$N(a\mu_1+b\mu_2, (a\sigma_1)^2+(b\sigma_2)^2)$となる。(正規分布の加法性)

- 日本人男性成人身長の分布 N(172.5 cm, $(5.5 \text{ cm})^2$)
- 日本人女性成人身長の分布 N(158.7 cm, $(5.3 \text{ cm})^2$)
  - H25年度 30~34歳 (総務省 統計局 http://www.e-stat.go.jp/SG1/estat/Xlsdl.do?sinfid=000027237853)
- 無作為に日本人成人男女2名を選び出すと、その身長の差の分布は
  $N(172.5\text{-}158.7 \text{ cm}, 5.5^2+(-5.3)^2 \text{ cm}^2)$
  $=N(13.8 \text{ cm}, 58.3 \text{ cm}^2)$
  $=N(13.8 \text{ cm}, (7.6 \text{ cm})^2)$
  - 平均 $\mu=13.8$ cm 標準偏差 $\sigma=(58.3)^{0.5}=7.6$ cm

正規分布の加法性

- $N(\mu_1,\sigma_1^2)$の分布のxと、$N(\mu_2,\sigma_2^2)$の分布のyがある時、ax+byの分布は$N(a\mu_1+b\mu_2, (a\sigma_1)^2+(b\sigma_2)^2)$となる。(正規分布の加法性)
- 日本人男性成人身長の分布 N(172.5 cm, (5.5 cm)^2)
- 日本人女性成人身長の分布 N(158.7 cm, (5.3 cm)^2)
  - H25年度 30~34歳 (総務省 統計局 http://www.e-stat.go.jp/SG1/estat/Xlsdl.do?sinfid=000027237853)





正規分布の加法性

- $N(\mu_1,\sigma_1{}^2)$の分布のxと、$N(\mu_2,\sigma_2{}^2)$の分布のyがある時、ax+byの分布は$N(a\mu_1+b\mu_2, (a\sigma_1)^2+(b\sigma_2)^2)$となる。(正規分布の加法性)

- 日本人男性成人身長の分布 N(172.5 cm, (5.5 cm)^2)
- 日本人女性成人身長の分布 N(158.7 cm, (5.3 cm)^2)
  - H25年度 30~34歳 (総務省 統計局 http://www.e-stat.go.jp/SG1/estat/XlsdI.do?sinfid=000027237853)
- 無作為に日本人成人男女2名を選ぶと、その身長の差の分布はN(13.8 cm, (7.6 cm)^2)
- もし世の中の男女カップルが身長差とは無作為に相手を選んでいるとしたら
  - 逆身長差カップル(女性の方が高い) は全体の3.44% (1/29)
    - (0-13.8)/7.6=-1.82σ
  - 高身長差カップル(男性の方が30 cm以上高い)は全体の1.66% (1/60)
    - (30-13.8)/7.6=2.13σ

中心極限定理・大数の法則

- **中心極限定理**
  - どんな確率分布であれ、たくさんのランダムな数の和の分布は、**正規分布**に近づく。
  - 例) サイコロ(一様分布)を100回振った和の分布は、正規分布に近づく。
  - この定理のため、正規分布は世の中のあちこちに現れる。

- **大数の法則**
  - サンプル数を増やしていくと、サンプルから計算した平均は**真の平均**に近づく。
  - 例) 日本人男性10人から計算した平均身長より、1000人から計算した平均身長の方が、日本人男性の真の平均身長に近い。
  - 当たり前。
  - ただし、正規分布の加法性と中心極限定理が大数の法則の成立を保証している。

興味があれば、こちらをどうぞ…

統計学入門 (基礎統計学 I)

東京大学教養学部統計学教室

東京大学出版

3024円